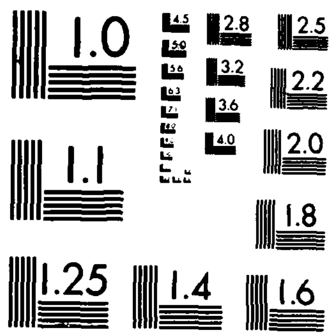
AD-A174 288 CORRELATION OF EXPERIMENTAL AND EXERCISE RESULTS(U) 1/1
DEFENSE SYSTEMS INC MCLEAN VA 30 MAY 86
DCA100-86-C-0004

UNCLASSIFIED F/G 17/2 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A174 288

DTIC FILE COPY

CORRELATION OF EXPERIMENTAL AND EXERCISE RESULTS

30 May 1986

Prepared For:

Contract DCA100-86-C-0004

Headquarters Effectiveness Evaluation
Defense Communications Agency
Washington, D.C. 20305

DSI Defense Systems, Inc.
7903 WESTPARK DRIVE
MCLEAN, VIRGINIA 22102
(703) 883-1000

86 11 19 014

1

CORRELATION OF EXPERIMENTAL

AND

EXERCISE RESULTS

30 May 1986

Prepared For:

Contract DCA100-86-C-0004

Headquarters Effectiveness Evaluation
Defense Communications Agency
Washington, D.C. 20305

Prepared By:

Defense Systems, Inc.
7903 Westpark Drive
McLean, Virginia 22102
(703) 883-1000

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A174 288

1471

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp Date Jun 30 1986

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b RESTRICTIVE MARKINGS: N/A

2a SECURITY CLASSIFICATION AUTHORITY: N/A

2b DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

3 DISTRIBUTION/AVAILABILITY OF REPORT: A - Unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S): N/A

5 MONITORING ORGANIZATION REPORT NUMBER(S):

6a NAME OF PERFORMING ORGANIZATION: Defense Systems, Incorporated

6b OFFICE SYMBOL (If applicable):

7a NAME OF MONITORING ORGANIZATION:

6c ADDRESS (City, State, and ZIP Code):
7903 Westpark Drive
McLean, Virginia 22102

7b ADDRESS (City, State, and ZIP Code):

8a NAME OF FUNDING/SPONSORING ORGANIZATION: Defense Communications Agency

8b OFFICE SYMBOL (If applicable):

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DCA100-86-C-0004

8c ADDRESS (City, State, and ZIP Code):
Defense-Wide Systems Directorate (A700)
Washington, DC 20305

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
		3	006

11 TITLE (Include Security Classification): Correlation of Experimental and Exercise Results (Unclassified)

12 PERSONAL AUTHOR(S): DSI C3I Team

13a TYPE OF REPORT: FINAL

13b TIME COVERED: FROM N/A TO

14 DATE OF REPORT (Year, Month, Day): 860530

15 PAGE COUNT: 49

16 SUPPLEMENTARY NOTATION: N/A

17 COSATI CODES

FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number): ~~N/A~~ C^2 : BFIT : NPS : HEAT

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This report addresses the relationships between recent experiments, exercises, and insights into the elements of command and control ~~(C2)~~ theory. It summarizes and substantiates insights arising from the application of the Headquarters Effectiveness Assessment Tool (HEAT) to experiments at the Naval Postgraduate School (NPS) and Battle Force In-Port Training (BFIT) exercises in the Second Fleet. It also addresses the relationships between experimental results and exercise results in general, and concludes with a set of guidelines for future research.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL: Deborah L. Therrien

22b TELEPHONE (Include Area Code): 703-883-1073

22c OFFICE SYMBOL:

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

CORRELATION OF EXPERIMENTAL
AND
EXERCISE RESULTS

Table of Contents

List of Figures

List of Tables

INTRODUCTION

This report addresses the relationships between recent experiments, exercises, and insights into the elements of command and control (C^2) theory. It summarizes and substantiates insights arising from the application of the Headquarters Effectiveness Assessment Tool (HEAT) to experiments at the Naval Postgraduate School (NPS) and Battle Force In-Port Training (BFIT) exercises in the Second Fleet. It also addresses the relationships between experimental results and exercise results in general, and concludes with a set of guidelines for future research.

Over the past three years, Defense Systems, Inc. (DSI) has supported the Defense Communications Agency (DCA) in a program to define, measure and identify determinants of C^2 effectiveness. This program proceeds along three parallel tracks--(a) the development of theory (identification of key concepts, specification of definitions, hypotheses linking different concepts, theoretical formulations, prediction of observable patterns, etc.); (b) the conduct of experiments and exercises in order to test hypotheses, generate new insights and validate the approach in the "real" world; and (c) the development of a knowledge base (some empirical parameters, some data and some insights) that relates theoretical development to empirical observation. This cycle (theory, empirical observation, theory) is the only useful route to building knowledge in a complex, poorly specified, and dynamic world such as military command and control.

Three sources of empirical data have been used to date in the DCA research program: historical data, observations of exercises, and laboratory experiments. These sources differ considerably in the realism of the observed activity, the degree of experimental control, and the repeatability of the results. The distinctions between them must therefore be kept

clear; but at the same time there needs to be an understanding of how the data from the different sources relate to one another, support one another, and contribute in their different ways to a single body of knowledge. The need for clarifying these relationships applies especially to the recent laboratory experiments at the NPS and the BFIT exercises, in which DCA and DSI have participated.

The laboratory experiments had among their objectives the test of propositions, formulated mainly in theoretical work but influenced also by observation of the BFIT exercises. Three successive yearly sets of experiments have addressed, respectively, the effects in a C^2 network of varying

- degrees of connectivity among nodes in the network;
- degrees of centrality in the C^2 structure; and
- the assigned role and specialization of a command.

This report begins by recapitulating the tested propositions and the degree to which each of them has been confirmed. New insights arising from the experiments, and from exercises conducted in the same time period, are also summarized.

The report continues with a discussion of the general relationships between experiments and exercises. As data sources for C^2 performance, experiments and exercises fall near the middle of a scale that ranges from mathematical models to histories of warfare:

- Mathematical models
- Computer simulation
- Experiments
- Wargames

- Command post exercises
- Field exercises
- Actual combat

The discussion of the relationships between these sources focuses on key variables that serve to quantify the differences between them. These include dependent variables descriptive of the outputs from a C^2 organization, of which HEAT measures of effectiveness and process quality are a prime example; and control variables, which describe the differences between the C^2 organizations, their capacity, and the context in which they operate. An attempt is made in this report to define additional control variables which describe, numerically, the position of a C^2 operation on the scale shown above, in terms of both realism and the degree of control available to an investigator (the two vary more or less inversely).

Comparison of exercise and experiment results is made, at least initially, in terms of all these variables. The report describes a series of steps which should lead to useful and valid conclusions from the comparison. The place of statistical techniques, as well as their limitations, is also discussed. The specific techniques used in the comparison of the DCA-observed experiments and exercises are given in detail.

Review of the three NPS experiments and two BFIT exercises has identified four HEAT measures and 26 control variables that can be computed for both groups. These are tabulated in this report, and the patterns of similarities and differences among the variables are pointed out. Statistically significant differences among the dependent variables are also identified. These differences, and the pattern of the pertinent control variables, suggest that the effects observed in the laboratory

experiments are "real" in the sense that they occur in the real world also. This conclusion applies, in particular, to the propositions that the experiments were designed to test.

The report concludes with a set of general guidelines for future research. Theoretical work conducted by DSI from 1982 onward, supplemented by experiments and exercises such as those addressed here, has produced an extensive list of C^2 research topics. This report recapitulates recommendations for the most suitable research vehicle (experiments, exercises, etc.) in each case.

INTEGRATED FINDINGS FROM EXPERIMENTS AND EXERCISES

DCA and DSI have supported and participated in a continuing series of C^2 laboratory experiments at the U.S. Naval Postgraduate School. To date, three sets of experiments have been completed and analyzed:

- Experiments in the effects of connectivity, conducted in November 1983;
- Experiments in the effects of headquarters centrality, conducted in October 1984;
- Experiments in the effects of role and specialization, conducted in November 1985.

All of these experiments had among their objectives the test of propositions, formulated either in theoretical work or from observations of earlier experiments and exercises. This section of the report recapitulates those propositions, identifies their sources and discusses the extent to which they have been validated or disproven. Insights arising from the experiment results and from concurrent exercises are summarized as well.

PROPOSITIONS TESTED

Reference 1 is a compendium of proposed research in C^2 theory and, more specifically, in determinants of headquarters effectiveness. It was part of a larger study (including also References 2 and 3) dealing with theater headquarters effectiveness, its measurement, and its determinants.

Among those determinants are the internal architecture of the headquarters, and more specifically its connectivity. Several propositions relating the internal structure of a headquarters to its effectiveness were derived from the theoretical work of Reference 3 and set forth in Reference 1. Of these, the following proposition was tested in the 1983 experiments:

A: Star structures work faster but less accurately than multiconnected structures.

This proposition is derived from Soviet theory and experiments reported in Reference 4; and can be elaborated, as it was for the 1983 experiments, as follows. (Propositions are labeled A1, A2, etc. for later reference.)

A1/A2: Fully-connected C^2 systems are slower than minimally-connected, highly-centralized C^2 systems. Intermediate connectivity is faster than full connectivity. This relationship--decision speed increases as connectivity lessens--holds for both creative and formatted decisions.

A3/A4: Fully connected C^2 systems make better--more correct--decisions than minimally-connected, highly-centralized systems. The correctness advantage of full connectivity is greater for creative decisions than formatted decisions.

The theory presented in Reference 3 included a discussion of the critical determinants of the best assignment of tasks to headquarters. This assignment involves a specification of the level of detail at which the headquarters seeks to control its forces. The level of detail can be described in terms of the headquarters' role type (on a scale ranging from control-free to interventionist) or of the type of directives issued (mission-specific, objective-specific, or order-specific). All of these distinctions correspond, in a broader sense, to the degree to which the headquarters centralizes or decentralizes (delegates) its decisionmaking.

The question of determining the most effective role type in a given set of circumstances was addressed further in Reference 5. The guidelines developed there were combined with predictions of classical control theory to produce the following propositions, which were tested in the 1984 experiments:

B1: Geographic C^2 is better than functional C^2 for discrete problems.

B2: Geographic C^2 is less vulnerable than functional C^2 to communications disturbance.

B3: Geographic C^2 is more vulnerable than functional C^2 to increased problem complexity.

B4/B5: Functional C^2 is better than geographic C^2 for complex problems if no disturbance occurs, but geographic C^2 invulnerability to communications disturbance offsets the initial functional superiority.

In this context, geographic C^2 is decentralized C^2; as stated in Reference 6, it corresponds to separate control systems, in that each C^2 node is on its own responsible for all warfare types within a geographic sector. Functional C^2, on the other hand, corresponds to a large, integrated, and internally co-ordinated control system, whose components happen to be geographically dispersed.

A practical example of different role types meeting different requirements was provided by a series of BFIT exercises conducted by the U.S. Second Fleet beginning in 1985. Specifically, in BFITs 2-85 and 1-86 the participating forces, constituting a naval battle force, operated in a command and control structure that incorporated both centralized and decentralized C^2. In this hybrid organization, three carrier battle groups (CVBGs) were organized geographically with two exceptions: strike warfare planning and execution, including strikes at sea, were centrally controlled by one of the CVBGs. The C^2 process in these exercises was evaluated using HEAT, as reported in References 7 and 8.

The same theoretical considerations that led to Propositions B1-B5 also led to the following propositions, which were tested in the 1985 experiments:

C1: Higher echelons are better at planning tasks.

C2: Lower echelons are better at battle management tasks.

C3: Propositions B1-B5 apply to hybrid organization as well as to functional organization, but the differences from geographic C^2 are less pronounced.

CURRENT STATUS OF PROPOSITIONS

The propositions tested to date in the NPS experiments have, on the whole, been confirmed, although sometimes only weakly. That is to say that the test statistics obtained from the experiments have conformed to patterns predicted by the propositions. Only one (Proposition A1, which asserts the superior speed of star structures for creative decisions) has been disproven outright. Table I summarizes these results. Detailed descriptions of the statistics, their derivation, and their implications are presented in References 6 and 9.

The unexpected contradiction of Proposition A1 is nevertheless consistent with the C^2 theory of References 5 and 10 upon which the most recent experiments have been based. Propositions A1-A4 were derived from Soviet findings reported in Reference 4. Related Soviet experiments, however, suggest that Soviet results may also reflect the preferences of the Soviet command culture, which generally favor centralized structures. The C^2 theory developed within the current program suggests that speed should improve, not so much with the pruning of the C^2 network, as with the ability of the network to adjust to traffic requirements.

Propositions B1-B5, although listed as "confirmed" in Table I, should not be regarded as "proven" in any logical sense. They are, rather, consistent with a coherent body of theory and supported to date by most of the experimental evidence. The 1985 experiments, however, have indicated that their validity may also depend on behavioral patterns among the C^2 personnel.

Table I. Status of Propositions

	Proposition	Exercise Where Tested	Status
A1:	Star structures are faster than multiconnected structures for creative decisions	1983	Disproven
A2:	Star structures are faster than multiconnected structures for formatted decisions	1983	Confirmed
A3:	Multiconnected structures make more correct decisions than star structures	1983	Confirmed
A4:	The correctness advantage of multiconnected structures is greater for creative decisions	1983	Confirmed
B1:	Geographic C^2 is better than functional C^2 for discrete problems	1984	Confirmed
B2:	Geographic C^2 is less vulnerable than functional C^2 to communications disturbance	1984	Confirmed
B3:	Geographic C^2 is more vulnerable than functional C^2 to increased problem complexity	1984	Confirmed
B4:	Functional C^2 is better than geographic C^2 for complex problems if communications are undisturbed	1984	Confirmed
B5:	Geographic C^2 invulnerability to communications disturbance offsets the functional superiority without disturbance	1984	Confirmed
C1:	Higher echelons will be better at planning tasks	1985	Uncertain
C2:	Lower echelons will be better at battle management tasks	1985	Uncertain
C3:	Propositions B1-B5 hold for hybrid vs. geographic C^2	1985	Confirmed (with qualifications)

Propositions C1-C3 can be regarded as confirmed only to the extent that the expected characteristics of hybrid C^2 were observed in one of the two groups of experimental subjects. The difference in behavior between the two groups permitted no conclusive findings as to the difference in C^2 at higher and lower echelons, and may limit the validity of the findings dealing with centrality.

The question of whether the propositions remain valid outside the laboratory is addressed in subsequent parts of this report.

INSIGHTS FROM EXPERIMENTS

Connectivity Experiments

Higher internal connectivity expedites creative decisions; a finding which contradicts earlier Soviet conclusions. Higher connectivity thus contributes to the efficiency of decision-making for all types of problems. Optimal decisionmaking will not necessarily use the full network of connections in all situations, but the potential for full connectivity should be provided.

Centrality Experiments

The collective findings regarding centralized and decentralized functions indicate that centralized ("functional") C^2 is the structure of choice only for complex problems with undisturbed communications. This implies that centralized structures should be used only when

- the complexity of the problem requires it; and
- undisturbed communications can be expected.

The vulnerability of centralized C^2 to communications disruption is already well recognized, and knowing the effects of such disruption should be of concern both to planners of new systems and to users of existing ones.

Role Experiments

Behavioral tendencies of a group may override the effects of echelon or centrality. The two groups of personnel in these experiments behaved in distinctly different manners. One group was reactive, i.e., it tended to react to enemy actions as they occurred; while the other was proactive and attempted to anticipate, predict, and avoid or allow for enemy actions. The proactive group tended toward central management even when the C^2 structure did not call for this. As a result, its performance did not vary with the nominal C^2 structure (hybrid or geographic) as would otherwise have been expected.

INSIGHTS FROM EXERCISES

Exercise Bold Eagle 84

In this joint exercise, HEAT was given one of its earliest applications with a view to investigating the potential usefulness of automatic data processing (ADP) in a joint task force headquarters (Reference 11). The investigation concluded that ADP can help to reduce

- low or irregular frequency of updates to monitored information;
- a focus on excessively narrow areas of interest; and
- communication delays leading to over-aged data.

None of these propositions have been tested in the Second Fleet exercises or the NPS experiments. However, increased use of automated decision aids by Navy battle staffs (the Joint Operational Tactical System or the Integrated Tactical Decision Aid) may permit observation ot some of these effects in the future.

Second Fleet Exercises

The hybrid C^2 structure used in these exercises is probably their most interesting feature. It permits centralization of key functions (strike and anti-surface warfare) in a dispersed force without imposing a completely centralized command structure. A similar hybrid was tested in some of the 1985 experiments at NPS, with somewhat inconclusive results because of the differences in group behavior among the experimental C^2 teams, mentioned earlier. The hybrid apparently has not been tested in conditions of severe communications disruption, although this would be of interest in the light of the 1984 centrality experiments.

SUMMARY

The 1983 and 1984 experiments at the NPS demonstrated the power of the laboratory approach. They showed that established propositions about the effects of connectivity and centrality could be quantitatively confirmed or (as appropriate) refuted. They also produced evidence of significant relationships among many of the C^2 parameters within each set of experiments. The most recent experiments were much less conclusive, apparently because of an unforeseen effect of human behavior. However, they suggest that this behavior is an appropriate subject for further investigation; and that the impact of a hybrid organization, and the analogies and differences between varying roles and varying structures, can be more precisely understood.

GENERAL RELATIONSHIPS OF EXPERIMENTS AND EXERCISES

This part of the report describes techniques available to describe the similarities and differences between experimental data, exercise data, and actual combat. It focuses on the key variables that serve to quantify the differences between these sources, and concludes with a description of statistical techniques for comparing the data. The actual comparison of the observed data is described in the next part of the report.

CONTROL VARIABLES

Representation of Combat

As data sources for C^2 performance, experiments and exercises fall near the middle of a scale that ranges from mathematical models to histories of warfare:

- Mathematical models
- Computer simulation
- Experiments
- Exercises
  - Wargames
  - Command post exercises
  - Field exercises
- Actual combat

All of these are, in some sense, representations of actual combat. Their position on this scale can be quantified by considering the degree to which they correctly represent the components of the C^2 cycle, and the degree to which they are under the control of the investigator.

This cycle is viewed here as it is in HEAT theory (Figure 1). The steps of the cycle are, of course, the headquarters processes, whose effectiveness is the subject of investigation. The notable differences among the representations listed above are seen in the way they represent not only the headquarters processes, but also the interfaces between the headquarters and its environment. Thus the input to the headquarters, the headquarters' internal processes, and their output to the environment serve as the indicators of correct representation.

To be more specific, we will categorize the representations of the processes and interfaces in two ways. They will be considered complete if all the significant measurable details of the real world (in combat) are measurable in the representation. They will be considered accurate if the quantities that are measurable correctly represent their real-world counterparts.

Field exercises, for example, are considered as close an approximation to combat as is reasonably available. The actions of the headquarters itself can be identical to those it would take in combat, given the same input. This identity also extends to the headquarters' output (messages and directives). The environment, however, is artificial, and produces not actual death and destruction but some simulation of it. The feedback to the headquarters reflects only this simulation. It may be complete (i.e., the headquarters may not be able to distinguish it from the real thing) but its accuracy is, at best, not tested. At worst, it may be completely unresponsive to the headquarters' action, as when the input to the headquarters is determined in advance and does not reflect the free play of an opponent. The loss of accuracy is, of course, offset by the fact that the environment is to some extent under control, so that the headquarters can be confronted with a situation of the investigator's choice.

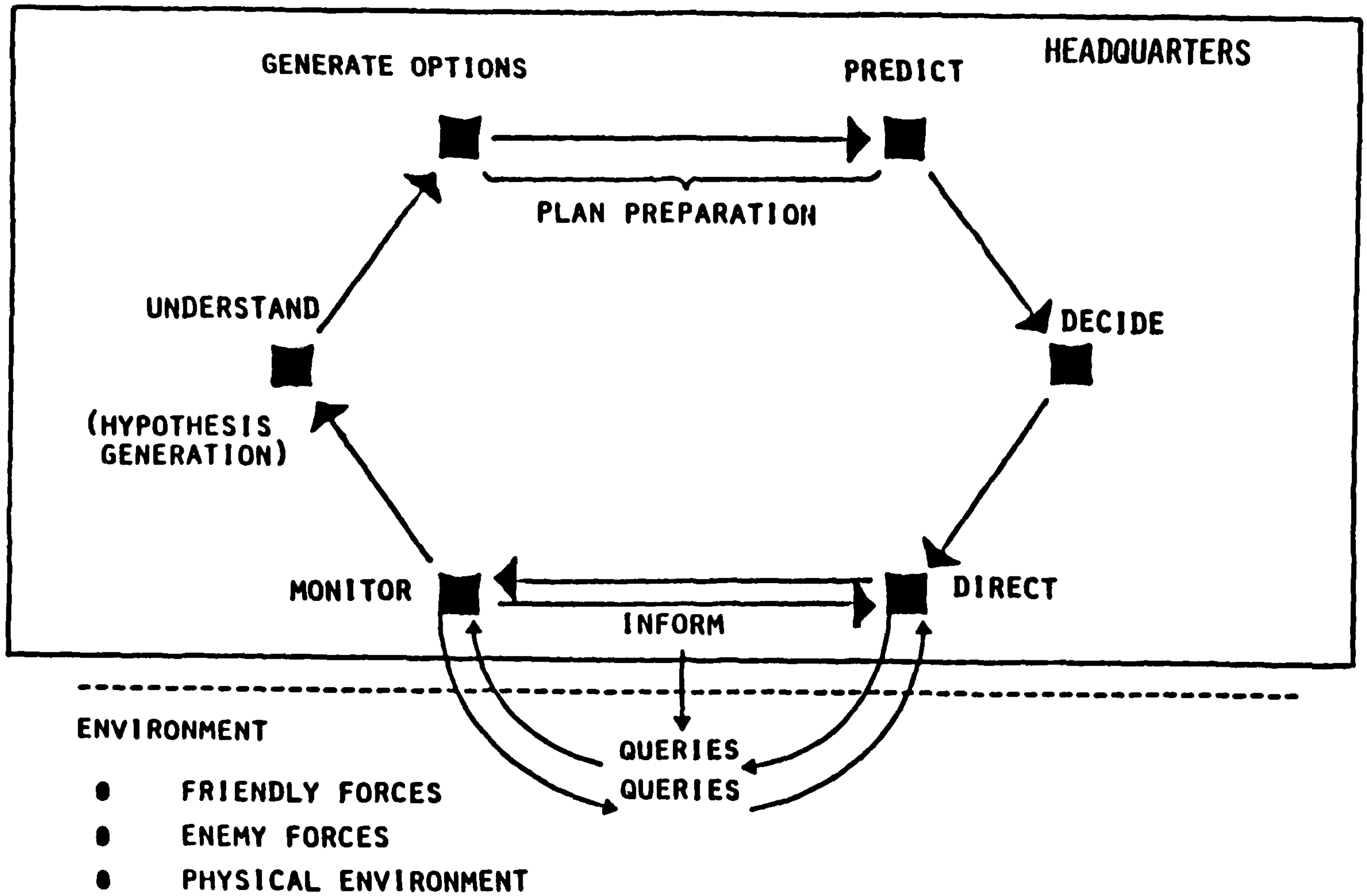


Figure 1. The Headquarters Cycle

In a typical command post exercise (CPX), the absence of interaction with actual subordinate forces leads to an additional loss of realism. The input to the headquarters is generated artificially; both its volume and its level of detail are reduced. The effects that the headquarters' actions have on the environment are no longer unpredictable, and may be prescribed by a script. Thus the representation of the environment and of the input which it generates is neither accurate nor complete. On the other hand, the investigator now has complete control of what that input will be.

A common feature of wargames and experiments is that the participants are no longer in their usual surroundings. Not only is their input artificial--again, it may be predetermined--but their output is likely to be in the form of instructions to a control team (or a machine) rather than the usual headquarters product. In this, the form of the output (and of the input) is constrained by the investigative technique and is no longer complete.

An additional artificiality which may or may not be present in experiments or wargames is the speeded-up clock. When the headquarters itself is no longer operating in real time, the representation of the headquarters processes is no longer accurate (at least in measures of time). Again, this artificiality is under the control of the investigator, who may indeed be interested in the effects of time constraints.

In computer simulations the parameters governing the (simulated) headquarters processes are entirely under the investigator's control. The outcome of the processes, and particularly of the planning and decision processes, no longer exhibit the unpredictability characteristic of real life. What remains is at most a complete (but not accurate) representation of the processes themselves.

At the extreme of the scale is the mathematical model, which normally avoids many details incorporated in a simulation and therefore represents headquarters processes only incompletely.

The above assessments are summarized in Table II, which also shows how the attributes of accuracy and completeness can lead to a numerical scale of relative realism. The scale here ranges from zero for mathematical models to six for actual combat.

Note that the representations discussed above and shown in Table II are merely typical examples. It does not follow, for instance, that every CPX will fall at four on the scale. A completely realistic message input might raise it to five; a working-hours-only schedule might lower it to three.

The relative degree of control by the investigator is summarized in the last column of Table II by crediting the representation with two points for full control of either the inputs or the headquarters processes, and one point for partial control. This scale ranges from zero for actual combat to four for mathematical models.

The degree of control is directly linked to the repeatability of results. When the investigator has control of inputs or processes, he is in a position to repeat the trial with the same parameters and to know that any variation in the results is due to uncontrolled factors. Repetition of trials in turn leads to greater precision of results (i.e., a smaller variance and a better estimate of their distribution). Thus a mathematical model, where control is complete, will always produce the same result unless, like the typical simulation, it incorporates a random process. At the other extreme, actual combat (as well as many exercises) is non-repeatable, and the data are imprecise in the sense that it cannot be known how

Table II. Representation of Combat

(1) Representation	(2) Environment (implies input)	(3) Input	(4) Processes	(5) Output	(6) Relative realism (number of entries, columns 3-5)	(7) Relative degree of control (input and processes)
Actual combat	CA	CA	CA	CA	6	0
Field exercise	C*	C*	CA	CA	5	1
Command post exercise	*	*	CA	CA	4	2
Wargame or experiment (real time)	*	*	CA	*A	3	2
Wargame or experiment (other)	*	*	C*	*A	2	3
Computer simulation	*	*	C*	*	1	3
Mathematical model	*	*	*	*	0	4

KEY:

C - Complete; all significant details measurable
A - Accurate; all measurable quantities identical with combat
* - Controllable by investigator

typical they are. The moderate degree of control associated with wargames or experiments lends itself to repeated trials, but the precision of the results is likely to be constrained by the cost of the trials as much as by unexplained variation.

Contextual Variables

The variables discussed in this section describe the context in which a headquarters operates, or in which headquarters performance measures are taken. They have also been called environmental or external variables, since they describe the environment in which the headquarters operates and which it seeks to control. Consistent with HEAT theory, the environment includes not only physical surroundings but also enemy forces (the principal objects of control) and friendly forces (through whom control is effected).

The variables described here are an extension of the contextual variables described in Reference 5.

Rate of change: the rate (per unit of time) of change in attributes of the environment that are of interest to the headquarters. C^2 theory suggests that this is the key to environmental influence on headquarters effectiveness. However, it is very difficult to measure, even in an artificial environment like that of the current experiments and exercises. Instead, some of the following variables can be used as indicators.

HEAT cycle frequency: the rate at which decision cycles are initiated. Although this variable is measured by observing the headquarters rather than the environment, it is a good indicator of the pace of battle.

Type of warfare: nuclear or conventional. Any qualitative differences between these types will probably also be reflected in other quantitative variables.

Number of units monitored, both friendly and enemy. This by itself is a rough measure of the complexity of the problem facing the headquarters (see below).

Problem complexity, defined as the number of different types of units monitored. Units are considered to be of distinct types insofar as they are controlled by different forms of orders (friendly units) or countered in different ways (enemy units). Present definitions are tentative and somewhat arbitrary. For the current exercises and experiments, the distinct types of units are:

- Friendly (in experiments)
  - Ships
    - Surface
    - Submarine
  - Aircraft
    - Combat
    - Support
- Friendly (in exercises)
  - Ships
  - Aircraft
- Enemy
  - ASUW
    - SAGs
    - Single ships
  - ASW
    - Missile-firing
    - Torpedo-firing
  - AAW
    - Missile-launching
    - Bomb-launching
  - Strike targets
    - Land
    - Sea

Problem subtlety: the number of meaningfully different futures consistent with the present evidence. The current experiments are designed with a predetermined level of subtlety; in the exercises to date, subtlety has been less of a concern except at the tactical level.

Defining Variables

The variables discussed in this section serve to define the headquarters in terms of its functions and its structure.

Echelon, or level of command. This variable alone can often serve to define the nature of a single headquarters and the distinctions between headquarters. The rest of the defining variables will generally be strongly correlated with the level of command.

Role type, or level of detail at which the headquarters attempts to exert control. For the present purpose of comparison, headquarters role types are categorized as mission-specific, objective-specific, or order-specific. Reference 5 contains a fuller description of these types. U.S. military headquarters can generally be described as objective-specific.

Number of subordinates, i.e., immediate subordinate commands to which the headquarters issues directives.

Number of nodes in the headquarters structure, including both internal and external nodes. An internal node is an organizational entity to whose activity a HEAT measure can be applied. An external node is an entity which either supplies an input to the headquarters or receives a headquarters output.

Links per node: the average number of links, to other nodes in the structure, possessed by each node.

Space distribution, describing the arrangements of nodes within the headquarters in terms of the form of communication which the arrangement permits. In this respect, a headquarters is categorized as

- integrated, when routine communications are conducted by face-to-face conversation;
- contiguous, when communications are predominantly by memorandum or telephone, although face-to-face communication is possible; and
- dissociated, when routine communications are conducted only by long-range means.

Connectivity, describing the degree of direct communication between nodes in their execution of the headquarters cycle. This is the percentage of direct links among the pairs of nodes that must communicate in performing the process steps in the cycle.

Capacity Variables

These variables describe the capacity of the headquarters to perform its assigned functions within its defined structure. Thus they serve to describe differences betwee.. otherwise identically defined headquarters, and possibly to explain differences in their performance. The resources that define capacity are personnel, automated data processing (ADP), linkages, information, and procedures. No descriptive variables have been identified for procedures as such, although Reference 5 provides a categorization of processing functions. The actual level of performance, of course, is a dependent variable (not a control variable) and measurable by HEAT.

Number of personnel, or "size": the number of people manning the headquarters.

Grade and specialty of personnel. For groups of personnel use the average (median) grade and most common specialty. For the headquarters as a whole, the personnel participating in the process steps Understand-Generate Options-Predict-Decide can be taken as representative.

Unit experience: the number of similar operations conducted previously by the same headquarters with the same personnel.

ADP usage: a listing of the headquarters process steps (including Inform) directly employing ADP.

ADP response time: the average delay in responding to a query.

Linkage reliability: the probability that a connection will exist between sender and receiver for the length of a transmission. This can be computed separately for each link; to describe the headquarters as a whole, the average reliability should be computed separately for external and internal links.

Linkage capacity, in terms of throughput rate. This is expressed as the densest kind of information that the link can handle (text, voice, data, or image, in increasing order). Capacity should be expressed separately for external and internal links. Internal links are often face-to-face, with capacity equivalent to "image".

Linkage medium, categorized simply as radio (line of sight), radio (beyond line of sight), wire, or none (no technical means). To describe the headquarters as a whole, give the most commonly used medium (separately for internal and external links).

Information accuracy, completeness, and timeliness appear to be the essential indicators of the quality of the information used by the headquarters. However, quantifying them depends on parameters and measurements that are difficult to obtain. Reference 5 suggests that accuracy be defined as the percentage of information within a desired accuracy window; completeness, as the percentage of inputs that specify all significant attributes; timeliness, as the percentage of information whose age is less than a desired value, or which is available when required. Clearly, these values cannot be computed without detailed supporting definitions and extensive, precise measurement (except perhaps for the last definition of "timeliness"). No attempt has been made to compute them for this report, and measuring them in any future operation will represent a significant effort.

Sampling density: the ratio of information presented to humans for input processing to total human input processing power. This variable also presents difficulty in measurement, and is not presented in this report.

DEPENDENT VARIABLES

As in earlier theoretic work, the HEAT measures of effectiveness and process quality are considered to be the primary dependent variables when headquarters and their performance are being compared. In the BFIT exercises dealt with by this report, all dependent variables measured were in fact HEAT measures, modified only slightly from the generic definitions in Reference 12. In the laboratory experiments, additional dependent variables were measured. These include "appropriateness" of headquarters actions, measures of communications activity, and unique HEAT-related measures, among

others. On the other hand, the constraints on C^2 in the laboratory limited the applicable generic HEAT measures to a handful. Comparisons between exercise and experimental results are of necessity confined to HEAT measures; the HEAT measures used in both cases are tabulated in the next part of the report.

COMPARING HEADQUARTERS OPERATIONS

Comparing different headquarters operations in which control variables and dependent variables have been measured involves several procedures. These include:

- Establishing whether the variables differ significantly;
- Determining patterns in the differences that have been identified; and
- Explaining those patterns in a way that permits prediction of other results.

Each of these will be discussed in turn.

Establishing significant differences in the observed variables involves asking, separately:

- Is there a significant difference between the results, i.e., the dependent variables?
- Is there a significant difference between the sets of control variables?

Differences between dependent variables are properly evaluated using statistical analysis. This is because there are enough unknown and unpredictable influences on the outcome of the headquarters process that the resulting measurements (e.g., HEAT scores) can be dealt with as random variables. Appropriate statistical techniques are described in the next section. No such technique can conclusively tell anyone

whether an observed difference in measurements is due to chance alone, or whether it arises from some underlying difference between the operations (e.g., different values of the control variables). This determination is a matter of judgment; judgment which can and should be supported by considerations apart from statistics. However, the statistical procedures can provide a precise index of the degree to which chance may have influenced the observed results.

Differences between control variables, on the other hand, are in no sense random. They are fixed, often planned, and (with some effort) measurable. Whether or not they are significant is again a matter of judgment; judgment to which statistical methods provide no support. There is, however, at least one technique that can simplify the task of making this judgment. That is the use of ordinal scales to describe any control variables that are otherwise expressed as numbers. As an example, links per node can be described as low, moderate or high, rather than as a precise number. Reference 5 suggests ordinal scales for many other control variables.

With significant differences identified, the next question is whether there is a pattern to them. This is a question to which there can (under certain circumstances) be a precise answer. The interrelated statistical techniques of correlation, regression, and analysis of variance are available to specify the degree to which one set of variables depends on another; or more precisely, to which changes in one set imply change in another. Thus it may be possible to describe quite precisely the mathematical relations between changes in the control variables and changes in the HEAT scores. It is also possible (just as it is when comparing results) to specify the degree to which chance alone may have produced these relations.

There are several limitations to these procedures which their user must recognize. First, the forms most commonly available deal only with linear relationships between variables. Second, to the degree that observed variations may be due to chance, the observed patterns may not necessarily justify the prediction of similar patterns in another comparison. Third, the probabilistic statements produced by these techniques implicitly assume that any unexplained variation is approximately normally distributed, which may not be the case. Finally (and consequently), a large number of samples (or observations, or comparisons) is needed to justify any confidence that the observed correlations are meaningful. For this reason, these techniques are usually applied only to large, carefully designed sets of experiments (such as the individual sets of laboratory experiments at the NPS). As the number of independent sources of data decreases, so does the meaningfulness of results. For example, if used to compare only two operations, these techniques will show perfect linear correlation (two points always form a straight line), but the level of confidence in any statistically supported judgment will be zero.

Statistically significant correlation between control variables and dependent variables does not necessarily mean that the difference in the control variables <u>caused</u> the change in the results. Nevertheless, the question of causality should be investigated, not only as an explanation of the observed pattern, but as a possible way to predicting effects in the future, leading (as necessary) to eventual improvements in C^2 performance generally. Causal connections, if they exist, will not be "provable" by statistical analysis. The evidence for their existence must draw on other supporting models, incorporating careful logical or mathematical reasoning. Such a model should then be scientifically testable, and the theoretical reasoning can help to suggest what statistical tests should be run to discover patterns of data inconsistent with its logic. Often, of course, the evidence does not rule out all competing explanations.

On the whole, however, investigation of the relationships between control variables and dependent variables is more likely to lead to insights and hypotheses, rather than to conclusions of cause and effect. Such insights may take the forms of propositions like those described in the first part of this report and, like them, be subject to further testing. Failing this, comparisons and observed correlations should at least be documented as potential sources for further investigation.

Methods of Statistical Comparison

The purpose of the techniques described in this section is to compare two samples of random variables and to provide the analyst with an indication of whether the differences observed between the samples are due solely to chance. All of these techniques can be viewed as statistical tests of the null hypothesis that both samples were drawn from identical "populations," i.e., that the probabilities governing the distribution of the observed variables were the same in both cases.

Each technique applies a measure of some sort to the two samples which describes the difference between them. This measure is itself a random variable (since it varies with the values in the samples) and is defined in such a way that, under the null hypothesis, its distribution is known. This permits the calculation of P, the probability that (under the null hypothesis) the difference would be as great as what was in fact measured. If P is fairly large, we have observed a result consistent with the null hypothesis and have no particular basis for saying that the two samples represent different distributions. If P is small, the null hypothesis may still be true; but in that case we have observed an improbable event, and we will be inclined to reject the null hypothesis. (We may do so at a "confidence level" of 1 - P. A commonly accepted confidence level is 95 percent, meaning that the null hypothesis is rejected whenever P is less than 0.05.)

Unlike many techniques in common use, the statistical comparisons described here do not assume that the two samples are normally distributed. Instead, they are "distribution-free," i.e., equally valid whatever the form of the underlying distribution. Normal distributions can indeed be expected when we are dealing with sums or averages of large samples; but sample sizes in these experiments and exercises are small, and the observed distributions of HEAT scores have been clearly <u>not</u> normal.

Four statistical techniques have been used in comparing NPS experiments and BFIT exercises:

- The Mann-Whitney "U" statistic;
- Likelihood-ratio tests on contingency tables;
- Run tests; and
- Median tests.

In all cases, they have been applied not to composite HEAT scores, but to the set of all instances of a particular HEAT measure being applied. Thus if the two samples (A, B) consist of scores

A: 3, 4, 6, 7
B: 1, 1, 5, 8, 15

we are comparing a sample of four values to a sample of five, not an average score of five to an average score of six.

<u>Mann-Whitney</u>. The U statistic is derived from the relative ranks of the two samples when they are combined and arranged in order. It is defined as the number of times B's precede A's in such an arrangement. (In the example above, $U = 10$.) Very large or very small values of U are unlikely under the null hypothesis. The distribution of U for small sample sizes can be found in tables (e.g., Reference 13), calculated

(Reference 14) or, for large sample sizes, approximated by a normal distribution. This is a relatively sensitive distribution-free test, and is appropriate whenever the data can be arranged in order. In the present application it was used on HEAT cycle times.

Contingency tables. When the scores take on binary values (e.g., correct or incorrect), the U statistic is insensitive to differences. In this case, the scores are arranged in a 2 x 2 table, for example:

	X	Y
Correct	2	0
Incorrect	4	5

The null hypothesis is, in effect, that the probability of a "correct" score is independent of whether we observe X or Y. The measure of the difference between the samples is the likelihood ratio (Reference 14). P is the probability that the likelihood ratio is as low as the observed value (or lower). This is the equivalent of Fisher's exact test (Reference 15). In the present application this test was used on all HEAT measures that are based on binary observations, i.e., correctness and completeness of understandings and predictions.

Run tests. When the two samples are combined and arranged in order, a "run" is a set of consecutive items from the same population. A low number of runs is unlikely under the null hypothesis; the distribution of this number can be calculated (Reference 14). This test was used to supplement the Mann-Whitney test since, unlike the latter, it is sensitive to differences in the shape of the distributions as well as in their location.

Median tests. This is a test on a 2 x 2 contingency table whose entries are the number of scores above and below the population median. In the first example above, this would be

	A	B
Above or equal	2	3
Below	2	2

This test serves as a quick substitute for the Mann-Whitney test (and was used here to confirm its results) but does not generally add any more information than the latter.

COMPARISON OF EXPERIMENTS AND EXERCISES

As of March 1986, HEAT has been used to measure performance in three sets of C^3 experiments at the Naval Postgraduate School and two Battle Force In-Port Training (BFIT) exercises in the U.S. Second Fleet. The variables describing these operations and their results are presented in Table III, beginning with control variables and ending with as many HEAT-related performance measurements as were applied to both groups.

All variables and measurements shown pertain not to a single headquarters, but to a network of 4-5 headquarters, which are the "nodes" of the network.

CONTROL VARIABLES

The degree to which the exercises and experiments approximated actual combat, and conversely the degree to which they were under control, is shown in the first two lines of Table III. The experiments correspond most closely to the non-real-time experiments of Table II, and the exercises, to the CPX. The primary reason for the differences between the two sets of numbers are that the BFITs were conducted in real time (with only slight artificialities) using actual headquarters staffs and facing relatively unstructured problems.

The tabulated contextual variables reflect the fact that the 1985 experiments were deliberately constructed to simulate a BFIT scenario. In the earlier experiments, where the focus of measurement was on planning ability, the monitored forces were smaller, while the subtlety of the problem was higher to permit more precise measurement of understandings. The HEAT

Table III. Comparison of NPS Experiments and Second Fleet BFITs

	NPS Experiments			Second Fleet BFITs	
	Connectivity (1983)	Centrality (1984)	Rate (1985)	2-85	1-86
Representation of Combat					
Degree of realism	2	2	2	4	4
Degree of control	3	3	3	2	2
Contextual Variables					
HEAT cycle frequency (per hour)	0.67	4.0	4.0	3.1	3.3
Type of warfare	Conventional	Conventional	Conventional	Conventional	Conventional
Friendly units monitored	31	84	154	236	398
Enemy units monitored	55	39	155	77	118
Problem complexity	12	12	12	10	10
Problem subtlety	9	18	4	4	4
Defining Variables					
Echelon	BF/BG	BF/BG	BF/BG	BF/BG	BF/BG
Role type	Objective-specific	Objective-specific	Objective-specific	Objective-specific	Objective-specific
Subordinates	31	84	154	28	67
External nodes	3	3	3	9	7
Internal nodes	4	4	5	5	5
Links per internal node	3.75-5.25	5.25	6.2	7	7
Space distribution	Dissociated	Dissociated	Dissociated	Dissociated	Dissociated
Connectivity	50-100%	100%	100%	100%	100%
Capacity Variables					
Personnel	4	12	13	40	40
Grade of planners	O3-O5	O3-O5	O3-O5	O3-O7	O3-O7
Unit experience	0	0	0	0-1	0-2
ADP usage	Monitor, Inform	Monitor, Inform	Monitor, Inform	Monitor	Monitor
External linkage reliability	100%	100%	100%	98+%	98+%
Internal linkage reliability	100%	96%	94%	98+%	98+%
External linkage capacity	Data	Data	Data	Image	Data
Internal linkage capacity	Text	Text	Text	Voice	Voice
External linkage medium	Wire	Wire	Wire	BLOS radio	BLOS radio
Internal linkage medium	Wire	Wire	Wire	BLOS radio	BLOS radio
Dependent Variables					
HEAT cycle time (directives) (min.)	4	--	--	23	11
Understanding correctness (enemy intent)	50%	65%	75%	100%	100%
Understanding completeness (enemy intent)	71%	--	78%	100%	67%
Prediction completeness	--	--	23%	34%	13%

cycle frequency is abnormally low in the first experiment, which did not call for separate planning at each node.

The defining variables, reflecting headquarters functions and structure, are similar if not identical in all five cases. The number of "subordinates" tends to be higher in the laboratory, where the headquarters must directly control all its units. Connectivity and links-per-node vary only in the 1983 experiment, where this topic was the principal subject of experimentation.

Capacity variables point up greater differences between the experiments and the exercises. The student teams forming the laboratory "headquarters" were of smaller size and lower average rank than their real-life counterparts, and were not in a position to gain long-term experience since each set of experiments involved a new class. ADP contributed to the monitoring process in all cases, through the Naval Tactical Data System (NTDS) or a simulation thereof; and the experiments also used electronic mail as their communications system. The "Image" entry for BFIT 2-85 reflects that exercise's use of Radar Video Recorder (RAVIR) inputs.

DEPENDENT VARIABLES

Relatively few dependent variables are common to the experiments and the exercises. In the exercises, all dependent variables measured were HEAT measures, modified only slightly from the generic definitions in Reference 12. In the experiments, the constraints on C^2 in the laboratory limited the applicable generic HEAT measures to a handful. The overlap between the two sets is shown in Table IV. As shown there, and also in Table III, comparisons are of necessity confined to four measures:

Table IV. Heat Measures in Exercises and Experiments

HEAT Measures Used in Exercises (Short Titles)[1]	Experiments in Which Used
Overall Plan Duration	None
Overall Plan Cycle Time	1983
Monitoring Accuracy	None
Monitoring Timeliness	None
Monitoring Querying	None
Monitoring Comparability	None
Understanding Duration	1983, 1984, 1985[2]
Understanding Formulation	1983[2]
Understanding Correctness at Implementation	None
Understanding Comparability	None
Option Coverage[3]	1985
Option Planners	None
Option Quantity	None
Prediction Duration	1985
Direction Contradiction	None
Direction Time from Decision	None
Direction Queries	None
Coordination Contradiction	None
Coordination Time from Decision	None
Coordination Queries	None
Coordination Time from Notification	None
Monitoring Report Accuracy	None
Monitoring Report Timeliness	None
Monitoring Report Adequacy	None
Monitoring Report Comparability	None
Understanding Report Duration	None
Understanding Report Comparability	None
Planning Report Duration	None
Information Timeliness	None
Information Queries	None

- - - - - - - -

Notes to Table IV:

1. The short title of the generic measure (from Reference 12) is shown. The titles were changed slightly in specific applications.

2. Among the categories to which this measure was applied, Enemy Intent was the only one appearing both in exercises and in experiments.

3. Insufficient samples in exercises for meaningful comparison.

- HEAT cycle times;
- Understanding correctness (enemy intent);
- Understanding completeness (enemy intent); and
- Prediction completeness.

A fifth measure, Options Coverage, was calculated in both cases, but the sample size in the exercises was too small to permit a meaningful comparison.

HEAT cycle times were recorded in the 1983 experiments as well as in the BFITs, but represent somewhat different processes. The laboratory values, with a median of four minutes, reflect adjustments to plans, based on the headquarters' exposure to successive small changes in the monitored situation, with planning time constrained by the artificialities of the experiment schedule. The BFIT values represent a mixture of minor adjustments and full planning cycles, including the generation of plans for complex evolutions such as strikes and sorties. The difference between the two BFIT medians of 23 and 11 minutes may not be significant; it fails the Mann-Whitney test even with as low a confidence level as 80 percent. The difference between the BFITs and the 1983 measurements is significant at better than 99 percent confidence.

The correctness of understandings can be compared between exercises and experiments only with respect to the category of enemy intent. It is not surprising that the highest scores occur in the BFITs, and in the 1985 experiments which used BFIT-related scenarios, since enemy intentions in the former were not intended to be ambiguous, except perhaps in tactical details. It is less clear why there was a steady increase in the experiment scores over the years, but differences in the scenarios may (upon closer examination) explain part of the pattern. There is no apparent correlation of understanding correctness with problem subtlety, but the differences in the scores are statistically significant (using Fisher's exact

test) at better than 90 percent confidence between the first and last experiment, and better than 99 percent confidence between the exercises and experiments as a whole.

The scores for completeness of understanding reflect the occasional failure to formulate such an understanding of enemy intent. Here the sample sizes from the BFITs are quite small (six from each) and according to Fisher's exact test there is no statistically significant difference between any two scores. The scores recorded here are consistent with a tendency to incomplete planning that has already been noted in reports on the BFITs (References 7 and 8).

These references also specifically point out the common failure to predict explicitly the outcome of the headquarters plan, although this is not an "official" HEAT measure. The only experiment (1985) to measure prediction completeness explicitly shows a score consistent with the BFITs. The difference between the two BFIT scores is not quite statistically significant at the 90 percent level of confidence, by Fisher's exact test.

PATTERNS OF DIFFERENCES

The control variables for the exercises and experiments are more similar than might have been expected. The conspicuous differences are in the degree of realism and in headquarters capacity. BFIT staffs were much larger, more experienced, and included higher-ranking officers. The remaining differences in capacity affect communications and do not, on the whole, favor either the experiments or the exercises. Differences in contextual and defining variables are few: exercise headquarters received inputs from more different sources, and had more friendly units to monitor, but fewer direct subordinates.

The dependent variables, i.e., the comparable HEAT measures, fall into two groups. The completeness of understandings and the completeness of predictions do not show a significant difference. On the other hand, HEAT cycle time is much longer in the BFITs than in the experiments, and the correctness of understandings is better in the BFITs.

The interesting thing about these two differences is that they do not arise from any of the conspicuous differences in control variables mentioned above. Instead, they arise directly from the degree of control over the scenario and over the schedule of operations. Cycle times are short in the experiments because they are constrained by the control team's direction and by the experiment schedule. Understandings are more often wrong in the experiments because the planners are presented with deliberately ambiguous evidence. (The challenge to planners in the exercises is not so much to interpret enemy intentions, as it is to produce complete plans and properly coordinate own-force operations.) The slight difference in the degree-of-control variable (one degree, for control of the headquarters processes) is reflected in the cycle times. The difference in correctness of understandings arises from the same degree of control (of inputs), exercised in different ways.

In summary, the exercise results correspond well to the experimental results, except where the results were directly affected by deliberate controlling actions. However, the results that can be thus compared are few in number and do not include the primary HEAT measure of effectiveness (plan viability).

IMPLICATIONS FOR REAL-WORLD OPERATIONS

The pattern of differences between exercise and experimental variables suggests that, with respect to the C^3 propositions tested to date, the laboratory results are a good indicator of real-world performance. The following facts support this statement:

- HEAT results were not significantly changed by the realism of the setting, except where they were directly affected by deliberate control of the exercise or experiment. However, the results used to test propositions were deliberately left free to vary.
- The propositions tested in the laboratory dealt with the internal organization and procedures within a network of headquarters. On the other hand, the control variables that describe real-world operations differ from those in Table III primarily in external matters, i.e.:
  - number of external nodes and links; and
  - degree of realism in the environment.

These facts are not conclusive, but suggestive; they support, although they cannot prove, the hypothesis that the effects observed in the laboratory experiments are "real" in the sense that they occur in the real world also.

GUIDELINES FOR FUTURE RESEARCH

This section describes appropriate vehicles for the research suggested by the experiments and exercises analyzed in the preceding section, as well as by earlier theoretical development. These guidelines are based on the previous characterization of experiments, exercises, other research vehicles, and the relationships between them. They address the question of where research should be conducted, given that results are sought which may eventually be of use in the real world.

The earliest and most comprehensive list of research topics developed in this program of C^2 theory and application is found in Reference 1. This list includes descriptions of the most suitable research vehicle in each case. These descriptions are summarized here in Table V. Newer propositions related to the experiments and exercises, and described earlier in this report, have also been incorporated in the table.

In many cases a research topic is most appropriately investigated initially at one level, with later validation of the results being conducted at a greater level of realism. Table V therefore shows appropriate research vehicles in two columns.

<u>Command and control laboratories</u> appear in Table V more often than any other recommended research vehicle. Indeed, a general procedure for choosing such a vehicle might usefully begin with serious consideration of laboratory experiments; after which, consideration can be given to methods offering greater control or greater realism, as appropriate. The

Table V. Guidelines For Research

	Source of Proposition (reference)	Appropriate Research Vehicles for Investigation	Validation
A. COMMAND AND CONTROL THEORIES			
1. Validity of HEAT			
THQ effectiveness is best measured in terms of impact on environment rather than of process quality	1		History
Timeliness of queries by a headquarters is a measure of the quality of monitoring	1	Laboratories Exercises	
Quantity of queries to a headquarters is a measure of headquarters ineffectiveness	1	Laboratories	
2. Optimum Task Assignment			
Scarcity of assets is critical to task assigment	1	Simulation Laboratories History	
Available decision time is critical to task assignment	1	Simulation Laboratories History	
Breadth of information is critical to task assignment	1	Simulation Laboratories Exercises	
Necessary span of coordinating authority is critical to task assignment	1	Laboratories Exercises History	
Optimum assignment may require different tasks assigned to different levels	7,8	Laboratories Exercises	

Table V. Guidelines For Research (Continued)

	Source of Proposition (reference)	Appropriate Research Vehicles for Investigation	Validation
A. COMMAND AND CONTROL THEORIES (Continued)			
3. Configuration of Theater Headquarters (THQ)			
The "minimum essential function" concept reassigns but does not reduce work	1	Simulation Laboratories Exercises	
Minimal command modules are infeasible over extended periods	1	Laboratories CPX	
Informal interstaff interactions affect THQ process speed and quality	1	Laboratories Exercises	
Separation of THQ elements does not impact performance	1	Laboratories Exercises	
Distributed headquarters require more personnel than unitary designs	1	Simulation Laboratories Exercises	
Commanders and senior staff need mobile command posts in distributed structures	1	Laboratories	
Distributed systems are slower and more accurate than unitary designs	1	Laboratories	CPX
B. INTERNAL HEADQUARTERS PROPOSITIONS			
1. Functional Separation			
Isolation of a function speeds its completion but delays coordination	1	Laboratories	
Isolation of a function decreases its speed	3	Laboratories	
2. Geographic Separation			
The more sites, the lower the headquarters process	1	Laboratories	
The more sites, the higher the quality of the headquarters process	1	Laboratories	

Table V. Guidelines For Research
(Continued)

	Source of Proposition (reference)	Appropriate Research Vehicles for Investigation	Validation
B. INTERNAL HEADQUARTERS PROPOSITIONS (Continued)			
2. Geographic Separation (Continued)			
The need for linkages rises faster then the number of sites	1	Laboratories	
For better performance: creative functions by co-located groups, structured functions by isolated individuals	1	Laboratories	
For better performance when communications are disturbed: all functions organized geographically	6	Laboratories	
3. Connectivity			
Hierarchical structures work faster than multiconnected structures	1	Laboratories	
Star structures work faster but less accurately than multi-connected structures	1	Laboratories	
Higher connectivity implies faster and more accurate solutions for both complex and structured problems	6	Laboratories	
4. Communications Limits on Headquarters Size			
Political interfaces are human-intensive	1	Laboratories	CPX History
Human input-output capacity sets the limit on information flow quantity and quality	1	Simulation Laboratories	CPX History
THQs are likely to experience overload in the digestion of information	1	Laboratories	CPX History

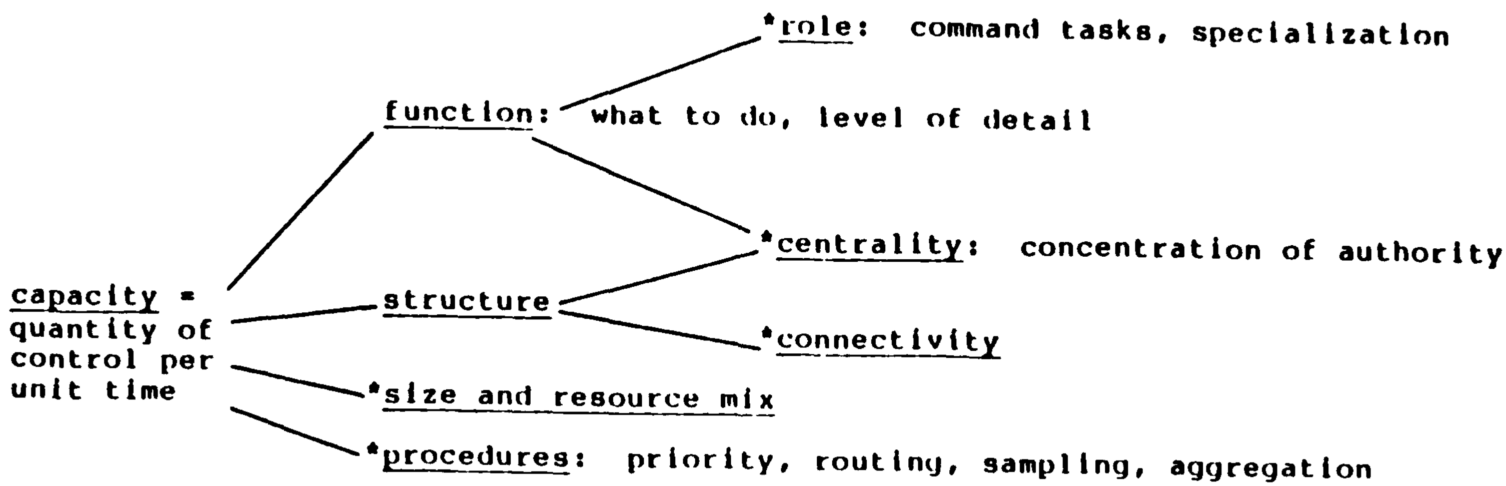
Table V. Guidelines For Research
(Continued)

	Source of Proposition (reference)	Appropriate Research Vehicles for Investigation	Validation
B. INTERNAL HEADQUARTERS PROPOSITIONS (Continued)			
4. Communications Limits on Headquarters Size (Continued)			
Headquarters effectiveness increases with linkage improvement only up to a certain point	1	Simulation	Laboratories
5. Convolution and Devolution			
Devolution of functions causes, at worst, an acceptable decrease in their effective performance	1	Simulation Laboratories	Exercises
The cost of devolving functions is less than or equal to the cost of convolving them	1	Simulation Laboratories	Exercises
6. Personnel Assigment			
Headquarters effectiveness increases with staff size with diminishing returns	1	Simulation History	Laboratories
Headquarters design needs to consider only the initial phase of war; migration of staff will approximately optimize personnel allocation	3	Simulation History	Laboratories
7. Automation			
Known sets of headquarters characteristics determine required numbers of different personnel types	1	History Experiments	History
ADP allows 3-1 to 4-1 reductions in personnel	1	History Experiments	History
ADP can reduce specific defects in the monitoring process	11	Laboratories	Exercises

laboratory setting offers the opportunity of observing actual human behavior, in circumstances largely under the experimenter's control. It allows variation of parameters and repetition of trials, neither of which is usually possible in more realistic surroundings; and it does so at a relatively low cost.

The current series of laboratory experiments at NPS is systematically examining primitive attributes of C^2 networks (i.e., those starred in Figure 2). These experiments have been directed at very high level issues consistent with HEAT's overall purpose--to discriminate between different levels of effectiveness. The results of the experiments to date suggest that this approach continues to be valid, with the next appropriate step being either to investigate an attribute such as procedures, or to remedy the inconclusive aspects of the role experiments. Other laboratory facilities (e.g., at the Army War College, the Air University, or the Naval Ocean Systems Center) may also be appropriate for such investigations. Laboratory settings can also be used, and most often are used, to explore more detailed issues--tradeoffs between humans and automated support, alterations of the pace of operations, etc. It is likely that future laboratory work involving HEAT will involve a richer mix of such causal variables and parameters.

Computer simulation may be the tool of choice where human behavior is not an issue, and the subject under investigation is relatively mechanistic. It is especially suitable when large sample sizes are needed to permit precise statistical analysis, or when investigation is made of system behavior over a timespan longer than can be represented more realistically. In the past, it has been very successfully applied in modeling systems dealing with data flow, transportation, communications, and analogous structures. As applied to command and control, it therefore is particularly good for addressing questions of efficiency, as opposed to effectiveness.



*candidate primitive attribute for early experiment

Figure 2. Using C^2 Network Attributes to Identify Early Experimental Topics

Mathematical modeling may appear as the underpinning for propositions to be evaluated, and as a source of new insights. However, as a primary means of investigation it plays no major role for the subjects addressed here. This situation may change as increased understanding of C^2 permits more accurate prediction of its operation. A mathematical model will be a useful part of this hierarchy of techniques only if its results can be translated into effects that are observable in combat, in experiments, and so forth. It must, in other words, deal with measurable quantities, so as to provide not merely logical support to a theory, but propositions that can be tested.

Exercises are typically the most realistic vehicle available for research into C^2 propositions. As such, they are often the most appropriate means for validating findings that have been substantiated in the laboratory or elsewhere. Field exercises permit observing the impact of C^2 on the environment, as long as the effects of actual combat are not at issue. They are expensive; but the C^2 application does not have to justify that expense, if the data can be collected (without interference) in an exercise conducted for training or other reasons. Command post exercises fall midway between field exercises and laboratory experiments in all these respects; they are most suitable for investigating questions of timeliness.

History of warfare provides the ultimate validation of any proposition. It is also the source of data and insights, and the two functions must be kept separate. In both functions, however, it has serious limits. As a means of validating findings, it is limited to whatever is on record: what is of interest to the present investigator may have been of no concern to the past historian. As a source of data, the same limitation applies. Nevertheless, it is the only way that the actual impact of C^2 in a combat environment can be observed.

The recommendations of research vehicles in Table V conform to the guidelines stated above. The appropriate methods for investigating new propositions can be derived directly from the considerations stated in the guidelines, or by analogy with related propositions already appearing in the table.

REFERENCES

1. Defense Systems, Inc., Theater Headquarters Effectiveness: Its Measurement and Relationship to Size, Structure, Functions and Linkages; Volume III: Approaches and Methods for the Validation of Determinants of Headquarters Effectiveness, 15 December 1982.

2. Defense Systems, Inc., Theater Headquarters Effectiveness: Its Measurement and Relationship to Size, Structure, Functions and Linkages; Volume I: Measures of Effectiveness and the Headquarters Assessment Tool, 15 December 1982.

3. Defense Systems, Inc., Theater Headquarters Effectiveness: Its Measurement and Relationship to Size, Structure, Functions and Linkages; Volume II: Design Consideration and Guidelines for Theater Headquarters Effectiveness, 15 December 1982.

4. Druzhinin, V.V., and D.S. Kontorov, Concept, Algorithm, Decision, USAF/GPO, 1972.

5. Defense Systems, Inc., Headquarters Effectiveness Program Summary, Task 002, 30 September 1983.

6. Defense Systems, Inc., C^2 Effectiveness Experiments: Estimating the Impact of C^2 Network Connectivity; Comparing Geographic and Functional C2 Organization, 15 August 1985.

7. Defense Systems, Inc., Second Fleet Battle Force In-Port Training 1-86 HEAT Analysis (U), Secret, 5 March 1986.

8. Defense Systems, Inc., Second Fleet Battle Force In-Port Training 2-85 HEAT Analysis (U), Secret, 28 May 1985.

9. Defense Systems, Inc., 1985 C^2 Effectiveness Experiments, 12 May 1986.

10. Defense Systems, Inc., Elements of C^2 Theory, 15 October 1985.

11. Defense Systems, Inc., Assessment of Headquarters Performance in Bold Eagle 84 (U), Secret, 31 May 1985.

12. Defense Systems, Inc., Headquarters Effectiveness Assessment Tool, User's Manual, 15 December 1985.

13. Bruning, James L., and B. L. Kintz, Computational Handbook of Statistics, Scott, Foresman, 1968.

14. Mood, Alexander M., and Franklin A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill, 1963.

15. Dixon, Wilfrid J., and Frank J. Massey, Jr., Introduction to Statistical Analysis, McGraw-Hill, 1969.

END

12-86

DTIC